DOCUMENT RESUME

ED 271 115 IR 051 552

AUTHOR Hardesty, Larry
TITLE Evaluation of Bibliographic Instruction.
PUB DATE 9 Apr 86
NOTE 49p.; Paper presented at the Library Instruction Roundtable of the Alabama Library Association (Montgomery, AL, April 9, 1986).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Libraries; Affective Measures; *Attitude Measures; *Cognitive Measurement; *Evaluation Methods; Formative Evaluation; Higher Education; Instructional Development; *Library Instruction; *Library Skills; Position Papers; Summative Evaluation

ABSTRACT

Arguing that there is a current tendency among librarians to talk more about the evaluation of bibliographic instruction than to actually do anything about it, this paper examines limitations of and considerations pertaining to evaluation and includes: (1) a brief discussion of the history of bibliographic instruction; (2) discussion of types of evaluation including formative and summative; (3) consideration of things most often evaluated, i.e., attitudes (affective domain) and skills (cognitive domain); (4) additional methods of evaluation including quantitative and qualitative; (5) issues of validity and reliability; and (6) conclusions and recommendations. Copies of overhead transparencies referred to in the text and a list of 50 references are included. (THC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED271115

EVALUATION OF BIBLIOGRAPHIC INSTRUCTION

by

Dr. Larry Hardesty
Director of Library Services
Eckerd College Library

Present to the
Alabama Library Instruction Roundtable
Alabama Library Association
Montgomery, Alabama

April 9, 1986

IR051552

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Larry Hardesty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TABLE OF CONTENTS

	Page
Introduction	1
Types of Evaluation	6
Formative Evaluation	6
Summative Evaluation	9
What Do We Evaluate	10
Attitude (Affective Domain)	10
Skills (Cognitive Domain)	12
How Do We Evaluate	17
Psychometric Characteristics	20
Validity	20
Reliability	23
Conclusions and Recommendations	25
References	27

TRANSPARENCIES

	Page
Formative Evaluation	6a
Instructional Development Model (Specific)	7a
Instructional Development Model (General)	7b
Revision/Recycle	7c
Tryout and Evaluation	7d
Selection of Media	8a
Summative Evaluation	9a
Attitudes (Affective Domain)	10a
Skills (Cognitive Domain)	13a
Quantitative	18a
Qualitative	18b
Validity	20a
Reliability	23a

EVALUATION OF BIBLIOGRAPHIC INSTRUCTION

Introduction

During the past two years many of you have enjoyed Florida's weather in December when you attended the Earlham College-Eckerd College Bibliographic Instruction Workshops. I thank you for inviting me here to have the opportunity to enjoy Alabama's April weather. A few years ago I introduced an article I had written on the evaluation of bibliographic instruction with a reference to a quote often erroneously attributed to Mark Twain on how "everybody talks about the weather, but nobody does anything about it."[1] Like the weather, we are still talking more about the evaluation of bibliographic instruction than we are doing anything about it. I hope that I can encourage some of you to do something about it. Since my background is from academic libraries, I will speak mainly from this perspective, but much of what I have to say is generalizable to other types of libraries.

To be fair, there have been some efforts to evaluate bibliographic instruction. Richard Werking reviewed and critiqued the handful of efforts he found in his article written in the Summer 1980 issue of Library Trends.[2] In addition, the Subcommittee on Evaluation, on which I served, of the Bibliographic Instruction Section of ACRL did publish a couple of years ago a handbook on Evaluating Bibliographic Instruction.[3] I recommend both Werking's article and the Handbook as introductory readings on the

evaluation of bibliographic instruction. Nevertheless, I still think there is a strong feeling that we are talking more about evalution of bibliographic instruction than actually doing it.

There are several reasons for our interest in the evaluation of bibliographic instruction and our lack of activity, and I will touch upon these reasons directly and indirectly during this talk. Sometimes, I suspect, we are afraid of what we will find out when we evaluate. I remember my first attempt at evaluation of bibliographic instruction. It was some dozen or so years ago soon after I graduated from library school and accepted a reference position in a state college in the Midwest. One day three graduate students who taught freshman composition came to the library to talk with the other reference librarian and me. Being conscientious individuals and relatively recent to their own undergraduate days, they wanted to incorporate a unit on how to use the library in their teaching. My colleague and I saw this as our "opportunity" to show what we could do. We worked with a media specialist to design a worksheet, an instruction manual, and a series of slides to accompany classroom lectures designed to teach the students how to use the library to write a research essay. Soon we had a couple of hundred students working diligently in the library and our librarian hearts were pounding proudly.

After the last worksheet was turned in and the last essay completed, we did as we thought we should. We went to talk with the graduate assistants about the "success" of the program. There is a saying among courtroom lawyers that "you should never ask a question you do not know the answer." However, being young and innocent of the ways of the world, I asked one of the graduate assistants, "Did the students write better

papers as a result of the exercise?" Now she knew the answer I wanted, but being equally young and naive, she gave me an honest answer, "NO." By her criteria the students spent too much time researching their topics in the library and not enough time writing their papers. She was, after all, a teacher of composition. I have not forgotten this lesson.

There long has been a perceived need by many, including most of us here today, for some way to enhance the educational role of the library. This perceived need goes much further back than the past ten to fifteen years. Some authors, such as Kenneth Brough, in The Scholar's Workshop,[4] trace early bibliographic instruction efforts back to the 1820s. Other authors, such as Mark Tucker[5] and Peter Hernon,[6] trace bibliographic instruction efforts back to the late 19th and early 20th century.

I suspect that the modern movement of bibliographic instruction was really started by Harvie Branscomb with the publication in 1940 of Teaching with Books.[7] In this landmark book, Branscomb examined a number of earlier studies and concluded that "it seems clear enough that undergraduates do not make very much use of the college or university book collection, not nearly so much as is ordinarily assumed."[8] This book had a tremendous impact at the time, and it continues to be timely in many ways. If you think you really have a new idea regarding bibliographic instruction, read Branscomb and you will be humbled. He anticipated many of the things we are doing and hope yet to do. Unfortunately, soon after the publication of Branscomb's book World War II diverted our attention for a few years and probably set back for another decade the evolution of bibliographic instruction.

We recognized during the late 1940s and early 1950s that a problem existed, but we really could not put our finger on it. In 1951, the eminent librarian Louis Round Wilson wrote "Although colleges spend a considerable proportion of their educational budgets for library materials and services, the contribution that the libraries make to further the educational program is less than what it should be."[9] It took Patricia Knapp in her classic article, "A Suggested Program of College Instruction in the Use of the Library", written in Library Quarterly in 1956 to first outline the foundations of modern bibliographic instruction.[10] She followed up her theory with experimentation and analysis in the Monteith College Project during the early 1960s.[11] By the late 1960s bibliographic instruction throughout the curriculum emerged with Evan Farber at Earlham College.[12] Within a few years later we had LOEX and the blossoming of the recent movement in bibliographic instruction.

In case you are wondering, there is a reason for this quick trip through the history of bibliographic instruction—other than I recently co-edited a book on the subject.[13] Despite several generations of effort and at least almost one full generation of the modern bibliographic instruction movement, bibliographic instruction has not been fully accepted in higher education. Gresham Riley, President of Colorado College, in speaking on the myths and realities of the academic library at the 3rd National ACRL Conference in 1984, stated that it is a reality that:

> Academic faculty are, for the most part, not predisposed as scholars to recognize and to acknowledge a legitimate educational role for the library and for librarians . . . they

> consider mastery of the library to be an easy task; and they believe extensive use of the library to be appropriate for only the brightest of students. For the majority of students, the library should serve as a study hall. Clearly, then, the library is not the heart of a college if one takes seriously the beliefs and practices of many faculty members.[14]

One result, to quote Robert Broadus, professor of library science at the University of North Carolina-Chapel Hill, is that "Use studies have pointed up the fact that in many libraries . . . there are miles of books that are not borrowed for years and years."[15]

In telling you this, I am, no doubt, dangerously close to preaching to the converted. Nevertheless, I want to point out why most of us are interested in evaluation of bibliographic instruction. We want to demonstrate its effectiveness in order to legitimize it and to make it acceptable. Through acceptance we can justify present support and gain additional support for our efforts. There are any number of programs and practices in higher education that can stand less scrutiny than bibliographic instruction, but over a period of years they have been accepted and have a relatively secure niche in the budgetary process. I will qualify this statement in the sense that given the present economic and demographic situation in higher education many program once unquestionably accepted are now being questioned—particularly by individuals and groups outside of higher education, such as governors and state legislators. Nevertheless, as a relatively new kid on the block even within our own libraries, we have to

work extra hard to prove our worth. The educational road to adoption is littered with promising innovations that never overcame the not inconsiderable inertia in higher education.

What shall we do? First, we must recognize that we are going to be evaluated whether we want to be or not. Second, we must provide a good evaluation if we are going to become part of the mainstream of education. To quote Edward Suchman, author of the book Evaluative Research, "All social institutions or subsystems, whether medical, educational, religious, economic, or political, are required to provide 'proof' of their legitimacy and effectiveness to justify society's support."[16] Third, we must recognize that there is as much witchcraft as there is science in evaluation. Suchman goes on to write that "both the demand for and the type of acceptable 'proof' will depend largely upon the nature of the relationship between the social institution and the public. In general, a balance will be struck between faith and fact."[17] Sometimes in education, facts may even endanger what faith the public has in us.

Types of Evaluation

Formative Evaluation

In the field of education, with which librarianship is closely allied, evaluation is customarily divided into two types: formative evaluation and summative evaluation. (overhead transparency one: formative evaluation) I suspect we are really more interested in summative evaluation than in formative, but let me talk about formative evaluation first. Formative evaluation is concerned with program development and is useful in making methods of instruction more effective. The main concern is helping the

FORMATIVE EVALUATION

PURPOSE : PROGRAM DEVELOPMENT

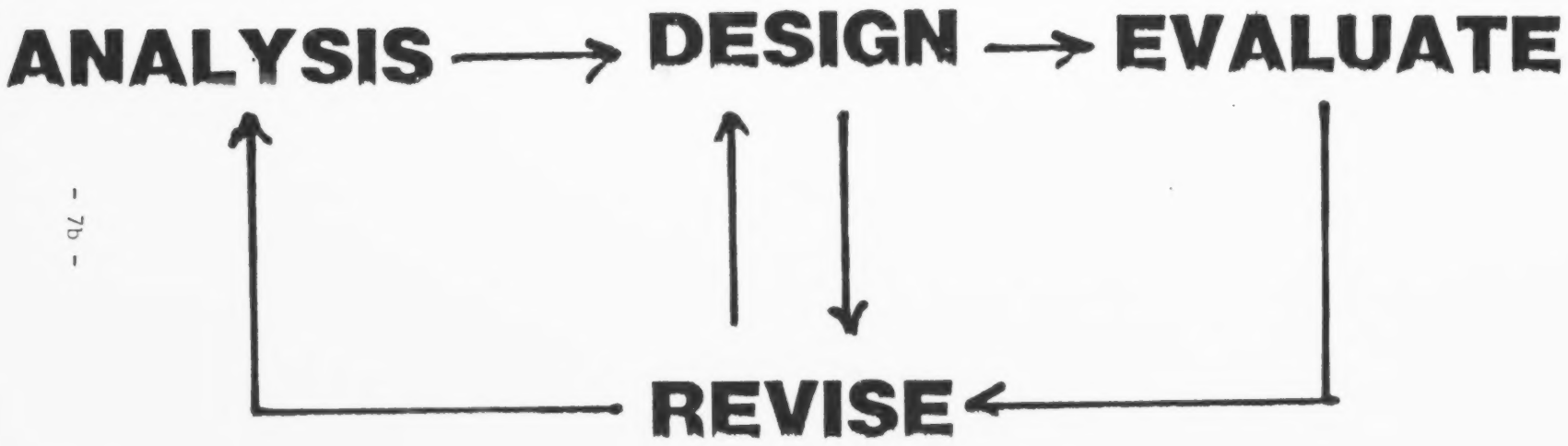
AUDIENCE: DEVELOPER

developer of a product or program, such as a worksheet, programmed booklet, slide/tape, videotape, lecture, etc., create something that is effective. Formative evaluation should be a consideration from the very beginning of the development process. There is a fairly elaborate instructional development model that has been created for this purpose. Let me give you one very brief example (overhead transparency two: instructional development model). Rather than try to go through this model let me do a couple of things. First, a few years ago I had a CLR-NEH grant to promote bibliographic instruction at DePauw University. As part of that grant I hired a doctoral student in instructional development (or Instructional Systems Technology as they call it at Indiana University). I had just finished a master's degree from that school in this field, and I thought this was a good opportunity to apply the model to a bibliographic instruction problem. We did an extensive literature search and documented our efforts in an article in the January 1980 issue of Drexel Library Quarterly.[18] I am, of course, biased, but I think it is a fairly complete but concise description of the application of instructional development principles to bibliographic instruction problems.

Second, let me simplify the model. It can be reduced to four steps (overhead transparency three: analysis, design, evaluate, and revise). There are some key things to remember regarding the revision loup (overhead transparency four: revision/recycle). Do not assume that the first draft will be the final version. Everything should be subject to revision. It is also useful to think of the formative evaluation process as a cone of widening experiences (overhead transparency five: tryout and evaluation). You can

INSTRUCTIONAL DEVELOPMENT MODEL

1. Needs Assessment
2. Solution Selection
3. Analysis of Task and Content
4. Development of Instructional Objectives
5. Selecting and Sequencing Instructional Strategies
6. Selection of Media
7. Development and Location of Necessary Resources
8. Tryout and Evaluation
9. Revision/Recycle
10. Social Aspects


ANALYSIS
DESIGN
EVALUATE
REVISE

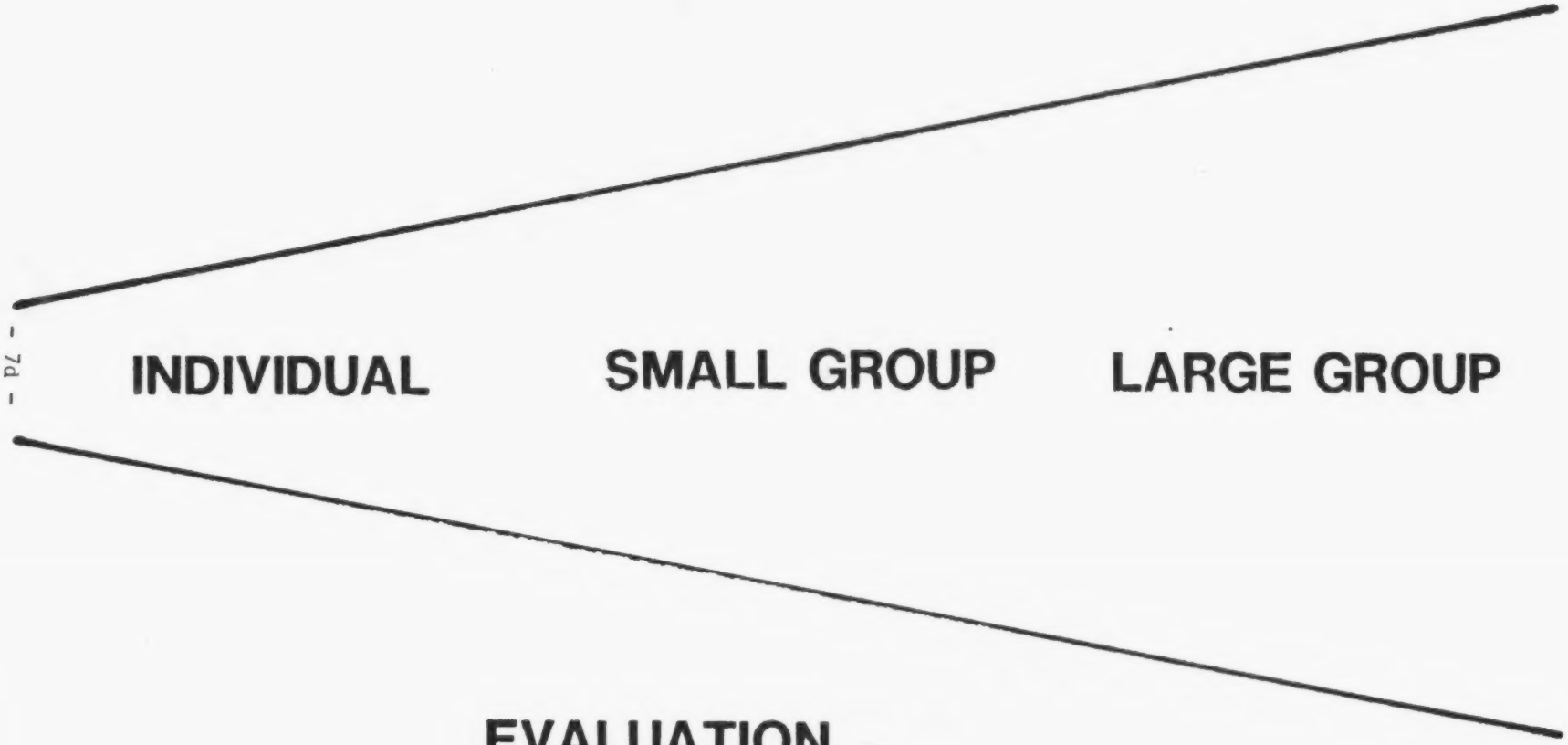
9. REVISION/RECYCLE

Don't Assume the First Draft

Will Be the Final Version,

Everything Should Be Subject to Revision

8. TRYOUT AND EVALUATION



start by simply having a few trusted student workers evaluate early drafts or versions. During each trial, students can be encouraged to note ambiguities, pacing difficulties, and other problems. After revision, the efforts can be expanded to small groups of students. Finally, you need to have the students toward whom the product is directed evaluate the program. Nevertheless, like Volkswagen, you should start small (overhead off).

There is one element of formative evaluation I want to dwell on here. Over the past few years there has been a growing number of articles in library journals comparing the effectiveness of various forms of media in bibliographic instruction. Finding the right medium for the right learning situation has been the "holy grail" among many audiovisual people for sometime. We should be careful not to allow ourselves to be sidetracked by this quest. Research has not given us data to allow sweeping generalizations to be made about media, such as "videotapes are better than slidetapes in providing an orientation to the physical organization of a library." This is not because of a lack of research in this area; there have been literally hundreds of studies conducted to compare the effectiveness of one medium with another. Nevertheless, few research studies have been able to take into account in reaching valid conclusions all the important variables. Given the present incomplete knowledge of media, librarians probably should base media selection on the issue of efficiency, considering the factors of flexibility, durability, costs, convenience, and availability. A particular useful guide to media selection has been developed by Robert Diamond (overhead transparency six: selection of media).[19]

6. SELECTION OF MEDIA

(Positive factors)

PROGRAMMED BCOKLETS

AUDIO CASSETTE

SLIDES

FILMSTRIP

PROGRAMMED BOOKLET WITH AUDIO AND/OR STILL VISUALS

TAPE FILMSTRIP

TAPE SLIDE

FILM CASSETTE SILENT

PROGRAMMED BOOKLET WITH FILM CASSETTE (SILENT)

FILM CASSETTE SOUND

VIDEO CASSETTE

Low

COST
(Per Station & Development)

High

High *FLEXIBILITY* *Low*

(Negative factors)

Time for completion - Internal
Content individualization

Summative Evaluation

Summative evaluation (overhead transparency seven: summative evaluation) deals with efforts to assess the overall effectiveness of a program and to gain additional support for the program. I was interested in summative evaluation when I asked the graduate student, "Did the students write better papers?" Summative evaluation reports are directed primarily toward those who set policy at various levels. I do not remember, but I suspect that I was all set to go back to the library director or the head of the English department with glowing reports of the "success" of the bibliographic instruction effort. In a strict sense, the summative evaluation should be at the end of the program and conducted by someone independent of the program. Both being and seeming objective are important if conclusions reached by summative evaluators are to be valid and believable.

Patricia Knapp's assessment of the Monteith College Project is a good example of summative evaluation—although she probably was more involved in program development than a good summative evaluator should be. I believe, however, there was a research analyst with training in sociology who did some of the analysis of the project. Another example of summative evaluation is the project I was involved with at DePauw University. A political scientist and a sociologist from DePauw worked with me on the quantitative part, which was published in two articles in College and Research Libraries.[20] A chemist from Earlham worked with me on the qualitative part, which was included in the final report to the Council on Library Resources and the National Endowment for the Humanities.

SUMMATIVE EVALUATION

PURPOSE : OVERALL EFFECTIVENESS

AUDIENCE : POLICYMAKER

What Do We Evaluate

Attitude (Affective Domain)

What do we evaluate? Although the stated primary purpose of bibliographic instruction programs is usually "to improve library use skills," we have a tremendous attitude problem on the part of students towards the library. Therefore, librarians frequently seek to evaluate the attitudes students have toward the library and librarians (overhead transparency eight: attitudes). Attitude changes, however, are difficult to measure. First, attitudes are implicit. That is, they are within the individual. An attitude cannot be seen, felt, touched, or observed in a direct fashion. We do assume, as expressed in most definitions of attitudes, that attitudes are a relatively enduring learned readiness to behave in a certain manner. We assume that attitudes have stability, relate to behavior, and can be changed through learning. In practice, any number of intervening variables can influence the acquisition and endurance of attitudes and the relationship between attitudes and behaviors.

I am convinced that attitude change is really the key to many bibliographic instruction problems. Many students either have negative attitudes from their own preconceptions regarding the library or they soon develop these attitudes from contact with their fellow students and the classroom faculty. They develop these negative attitudes for several reasons.

First, students have a natural human tendency to rationalize their own lack of knowledge and planning and to project these shortcomings on to something else—in this case, the library. Both classroom faculty and

ATTITUDES (AFFECTIVE DOMAIN)

1. IMPLICIT
2. LEARNED
3. RELATED TO BEHAVIOR

librarians often encourage students to develop unrealistic expectations regarding the rationale for the use of the library and the availability of library resources. Unrealistically high expectations inevitably result in frustrations and the development of negative attitudes toward the library.

I think this is an important point, and I want to elaborate on it. Human inquiry may be infinite but the library budget is not. Therefore, we need to match the level of inquiry we expect of our students with the availability of library resources. We, of course, can and should seek adequate funding for our libraries. This approach, however, has obvious limitations. A second approach is to reduce the level of inquiry we require of our students. At first glance this may seem like heresy at instituitons that encourage the free exploration of ideas. We, however, can limit the exploration of ideas and still accomplish the missions of our institutions. Particularly at the undergraduate level, students should be carefully directed in their library research. Their task is not "to push back the frontiers of knowledge," but "to discover the location of the frontiers of knowledge." With this in mind, I think we can establish some limits to the size if the library beyond simply the number of books we can purhase with the money we manage to get from the administration.

As a doctoral student seeking a dissertation topic, I received the same advice from several experienced researchers. They told me that there are any number of interesting questions that need to be explored and to be answered. A major challenge in any research, however, is to pick a topic that is explorable and answerable within available resources. If this is true for doctoral students and experienced researchers, it certainly is true for

undergraduate students. We need to convey to undergraduate students more realistic expectations of their research and thus reduce their often unrealistic demands on the library. Nevertheless, a change of attitude on the part of many students, classroom faculty, and librarians is necessary for the acceptance of this idea.

There may be an attitude construct known as "library anxiety." Connie Mellon discussed this construct in a recent issue of College and Research Libraries. Mellon proposed that students generally feel that (1) their own library-use skills are inadequate while the skills of other students are adequate; (2) the inadequacy is shameful and should be hidden; and (3) the inadequacy would be revealed by asking questions.[21] Obviously, we should be concerned and deal with library-related "anxieties" in our bibliographic instruction programs. I want to talk more about attitude when I discuss evaluation techniques and issues of validity and reliability. (overhead transparency off)

Skills (Cognitive Domain)

Generally, we have concentrated on the evaluation of librar skill development, or, that is, cognitive skills related to the library. These skills deal with recall or recognition of knowledge and the development of intellectual abilities and skills.[22] Bloomfield, in Lubans work Educating the Library User analyzed some sixteen library-use tests, about half of them dating from the 1930s. Bloomfield concluded that:

> The design of both library skills tests and courses tend to reflect what may be confused understanding on the part of test designers of how users need to use a library. Before

> we can design what are effective learning experiences in terms of either courses or tests, we are going to have to determine our library skills objectives and the needs of of students as prescribed by teachers with more understanding than we have had in the past.[23]

That is, before we evaluate library skills, we have to decide what they are.

Tom Kirk has provided a "taxonomy" of the evaluation of bibliographic instruction in a paper given at a LOEX conference and published in the proceedings Planning and Developing a Library Instruction Program in 1975.[24] Kirk suggested in addition to attitudes that we look at content, process, and product in the evaluation of bibliographic instruction programs. This approach is similar to efforts to evaluate the teaching of writing skills. Not only do we want to look at the final product, but we want to examine the development of the final product.

Often we try to analyze after the fact the process through which students use the library. Abraham Kaplan termed the result of such an analysis "reconstructed logic."[25] One caution is that reconstructed logic does not necessarily describe the logic actually used in the research process. In reconstructing the process used to find information, reseachers often over the role of intuition, education guess, and related intangibles.

A more widely-used approach to cognitive skills is Bloom's Taxonomy of Educational Objectives: Cognitive Domain.[26] Bloom and his colleagues organized cognitive skills into six major classes (overhead transparency nine: Bloom's Taxonomy). I will only briefly review them here. Knowledge is simply the "recall of specific and isolable bits of information."[27] For

SKILLS (COGNITIVE DOMAIN)

BLOOM'S TAXONOMY

1. KNOWLEDGE
2. COMPREHENSION
3. APPLICATION
4. ANALYSIS
5. SYNTHESIS
6. EVALUATION

example, where on the catalog card are "tracings" or "where in the library are the bound periodicals located?" These are examples of questions often asked on skills tests. I contend that they do not tap any sophisticated understanding of the library. Given enough bananas, I can train apes to respond to them correctly.

Comprehension relates to the ability of a student to know what is being communicated and to be able to make some use of the materials or ideas contained.[28] A library skills example might be the ability of a student to comprehend a Readers' Guide citation, or, a better example, to comprehend a Humanities Index citation after instruction on how to use Readers' Guide. Again, this is an elementary example.

Application, as Bloom defined the term, is the actual application of an ability.[29] A paper-and-pencil test may measure, for example, that a student can use an abstraction appropriate to a new situation. The test can be said to measure comprehension. However, if we use, for example, observation, journals, or some other method to determine if a student will use an abstraction correctly, then application is demonstrated.

Analysis is a more advanced level than the skills of comprehension and application. Analysis emphasizes the breakdown of material into its constituent parts and detection of the relationships of the parts and of the way they are organized.[30] Suddenly we are involved in some very sophisticated skills. We can, with a modest effort, teach a student how to locate a book on a particular topic in the library. It is much more difficult to teach and to evaluate the success of teaching a student how to analyze the organization of the literature of a particular discipline and how to

determine the interrelationships among, for example, books, periodicals, government documents, and other library resources in the search for information. Taken a step further, we can speak of analysis as determining the meaning of a communication. Once the student has found the book can he or she read it with any understanding? Where do library skills leave off and reading comprehension skills begin or is there a separation?

Synthesis is the putting together of elements and parts so as to form a whole.[31] Given the variety of resources available in the library, can the student not only find them and read them with understanding, can he or she also integrate the ideas found into a well-developed oral or written communication? Remember the question I asked of the graduate assistant? "Did the students write better papers?" They may have applied skills of knowledge and comprehension successfully and even analyzed the information found, but the program was not evaluated as a success by the graduate assistant because the students did not adequately synthesize the information found.

We run into all kinds of problems in this area. Sometimes the inability to synthesize information adequately results in plagiarism—which is not necessarily defined the same way among the various disciplines. Sometimes sophisticated synthesis is simply beyond the intellectual development of some students at the lower levels of college. William Perry, a Harvard psychologist, has written a book on the Forms of Intellectual and Ethical Development in the College Years in which he outlined the various stages students go through in their intellectual development.[32] He found through extensive interviews of Harvard under-

graduate students that students have various levels of intellectual maturity. They do not necessarily enter college at stage one and progress directly through the stages, but for the sake of simplicity let me outline the stages as such.

In stage or position one the students sees the world in polar terms of we-right-good vs. other-wrong-bad. There are right answers to everything, which are known to the teachers whose responsibility is to convey them to the students. In position two, the student perceives a diversity of opinion and uncertainty but accounts for the confusion in poorly qualified teachers (authorities) or as mere exercises set by the teachers, i.e. "so we can learn to find the answer ourselves." In position three the student accepts diversity and uncertainty as legitimate but only temporary because the teachers have not found the answer yet.[33]

Perry went on to describe a total of nine stages. Nevertheless, you can imagine the confusion in the minds of students in the earlier stages of intellectual development when they are assigned to go to the library to analyze the differing views regarding a particular phenomena, to synthesize these views, and to reach conclusions. I suspect that many students are simply overwhelmed and bewildered by the diversity of answers they find in the library.

Their inability to comprehend the diversity of viewpoints may be why some student, for example, do not want to use journal literature. They seem to want that one book which has the answer in it. In fact, it often seems that some students do think somewhere in the library there is one

large book with all the answers to their questions, and all the librarians have to do is show them where that one book is located. We, of course, maintain our job security by not showing them that one "magical" books.

Finally, the most sophisticated cognitive skill is evaluation, which is the making of judgments about the value, for some purpose, of ideas, works, solutions, methods, etc.[34] Evaluation sometimes is merely an expression of opinion based on utility. For example, a freshman student may give a book a good evaluation because it was short and easy to read. Generally, evaluation refers to judgments in terms of external criteria. For example, how do the ideas presented in a book relate to major theories and facts, are the ideas consistent, accurate, carefully written, etc.?

When we refer to library skills, I think it is very apparent that we very quickly move beyond the lower-order cognitive skills of recall and comprehension and involve the higher-order cognitive skills of analysis, synthesis, and evaluation. Not only are these skills harder to teach, they are also hard to evaluate. Nevertheless, if we do not consider them, we are not going have successfull bibliographic instruction programs. (overhead transparency off)

How Do We Evaluate

How do we evaluate library-related skills and attitudes? Methods of evaluation are generally divided into two categories: quantitative and qualitative or hard and soft. There are some topics, such religion and politics, that are hazardous to the health of friendships. Methodology issues should be on this list. Several disciplines, such as history, sociology, and psychology, are quite divided on this issue.

First, let us look at quantitative or hard evaluation methodology (overhead transparency ten: quantitative evaluation). One author said it is "regarded as rigorous by its friends and trivial by its enemies."[35] Generally, hard evaluation involves (1) a research design capable of discerning causal relationships; (2) the collection of data that are objective, reliable, and valid; and (3) the analysis of these data by sophisticated statistical techniques.[36] Use of paper-and-pencil tests are usually considered hard evaluations, and probably as good as any bibliographic instruction examples are two published studies I did at DePauw University.[37] For a critique of quantitative evaluation of bibliographic instruction you should read Richard Werking's article in the summer 1980 issue of Library Trends.[38]

Qualitative or soft evaluation methodology (overhead transparency eleven: qualitative evaluation) usually involves (1) a research design that, at most, can point to correlationships; (2) data that are subjective and judgmental; and (3) an absence of sophisticated statistical analyses of the data.[39] Case studies and use of interviews can be considered soft evaluation. While the characteristics I have listed of soft evaluations appear to emphasize their limitations, probably by far the most sophisticated and most useful evaluation of a bibliographic instruction program is a soft evaluation done by Patricia Knapp of the Monteith College program.[40]

Proponents of hard evaluation point out that soft evaluations are not easily subjected to replication and are often highly subjective. Proponents of soft evaluation point out that hard evaluators miss the essence of a

QUANITATIVE (HARD)

1. CASUAL
2. OBJECTIVE , RELIABLE , VALID
3. STATISTICS

QUALITATIVE (SOFT)

1. CORRELATIONAL
2. SUBJECTIVE AND JUDGMENTAL
3. LACK OF STATISTICS

program's impact, ignore variables that objective measures cannot assess adequately, and are mechanistic.[41] Werking in his article expressed this latter view when he wrote that "an increase in the average number of correct test answers from 12.2 to 14.9 on a scale of 20 would not convince all administrators that bibliographic instruction was ipso facto worth the thousands of dollars it was consuming, whether the gain was statistically significant or not."[42] Obviously there are few things on which one can convience all administrators. Nevertheless, administrators often commit much more than a few thousand dollars to programs with less demonstrated worth than provided by the skills test to which Werking referred. Sometimes administrators support programs despite negative evaluations. If you do not believe me, you should read some of the literature on the evaluations of the Head Start program.

A great deal of the success of evaluation depends on the values of those to whom the evaluation is directed. This consideration is worthy of a lot more discussion than I can give here. Nevertheless, comparisons of different methods of evaluation is sometimes like comparisons of the effectiveness of different media. A well-developed videotape will, of course, usually prove to be more effective than a poorly conceived slide-tape (or whatever other kind of media comparison you want to make). However, the differences in effectiveness quickly disappear when one applies equal funding and effort in their development. From all this controversy, one author concluded that there is no such thing as hard or soft evaluation--"only hard or soft evaluators."[43] (overhead transparency off).

Psychometric Characteristics

Whatever method of evaluation you use, you need to consider issues of validity and reliability—to which critics of each method really are referring. The Standards for Educational and Psychological Testing is the source for this information, particularly for quantitative tests.[44] It is published by the American Psychological Association, along with a couple of other professional societies. The most recent edition (1985) is a hundred page booklet that now cost almost thirty dollars. The previous edition is seventy-six pages and only cost about eight dollars.

Validity

What do we mean by validity? (overhead transparency twelve: validity) Basically, does the method used measure what it purports to measure. If it is a library skills test, does it really measure library skills? If we use structured interviews to measure changes in "library anxiety," are we really measuring "library anxiety"?

There are several components of validity. First, we need to be concerned with construct validity. If a measure has construct validity, then it actually measures the construct of interest. The problem is that we are not always sure what the construct is. Is there really something that can be termed "library anxiety," "reading readiness," or "clerical aptitude"? A construct is not just a label; it is a dimension understood or inferred from its network of interrelationships. If we cannot define it and distinguish it from other constructs, how can we measure it? Do library cognitive skills differ from reading skills, or, when we attempt to measure library skills are we simply measuring general reading skills or verbal ability? A good

VALIDITY

DOES IT MEASURE

WHAT IT PURPORTS TO MEASURE

example of a problem with construct validity is the call for evaluation of writing skills when we are not sure what these skills are. If a measure has construct validity, we should be able to infer that a high score means, for example, a high degree of "library anxiety" and a low score means a low degree of "library anxiety."

A second component of validity is content validity. Content validity depends on the adequacy with which a specific domain is sampled. For example, we may develop a paper-and-pencil test of library skills and only sample knowledge skills and never tap analysis, synthesis and evaluation skills. In this case, our test does not have content validity—if library skills are defined as the ability to analyze, synthesize and evaluate library materials.

Criterion-related validity refers to inferences that can be made from, for example, test scores, interviews, essays, etc., to an individual's most probable standing on some other variable called a criterion. If a student scores highly on a library skills test does that mean he or she actually will be able to make good use of available library resources? Or, better yet, does a low score mean that the student cannot make good use of the library? A student may not correctly identify the function of Readers' Guide on a test but may know how to use the set of green books in the index area that lists magazine articles in the library. A student may express a positive attitude toward the library in an interview, but because of peer pressure may not behave positively towards the library. There are all kinds of intervening variables that influence criterion-related validity.

Criterion-related validity can be referred to as concurrent validity if the criterion is measured at the same time or predictive validity if the criterion measure is given later.

I do not get a commission from the sale of the handbook on Evaluating Bibliographic Instruction but Bonnie Gratch has a nice section it on external and internal threats to validity.[45] She made good use of examples involving bibliographic instruction. Suppose you want to evaluate the results of bibliographic instruction on a group of students. You separate them into experimental and control groups and give them a valid and reliable test as freshmen and then you re-test them as seniors. Several things can happen to threaten the validity of the results. Maybe the less capable students with more negative attitudes toward the library and education in general drop out or transfer from school. Perhaps simply the maturation process resulted in more positive attitudes toward the library. Possibly the pretest sensitized members of the control group regarding their lack of library skills. Aware of their lack of knowledge, they somehow set about to correct this deficiency. These are all threats to internal validity. Any number of variables could account for the change or lack of change in the control and experimental groups.

Often we want to generalize our evaluations to other groups. Obviously we often want to make claims that our bibliographic instruction programs are successful and this success can be replicated with other students. What are some of the threats to external validity? Maybe we are able to interview or survey only students of cooperative professors. These students may be more sensitized to library issues, may be more

encouraged to have positive attitudes toward the library and librarians, and may be more reluctant to express negative attitudes than are other students. Conclusions reached from this group may not be valid for other groups of students.

Bonnie Gratch concluded that the traditional pretest-post-test with an experimental and control group is an extremely weak "pseudo-experimental" design.[46] I think that judgment is a bit harsh, but I suspect it is a reaction to those who claim too much from this methodology. Every methodology will have some threats to external and internal validity, whether it is quantitative or qualitative. If we waited until we had the perfect methodology to do evaluation, we would never do it. We simply must try to take into account and control these threats.

Reliability

The partner to validity is reliability (overhead transparency thirteen: reliability). Validity is the more important, but reliability is a necessary condition for validity. A reliable measure is one that provides consistent and stable indications of the characteristic being investigated. This can be an issue in the evaluation of bibliographic instruction in several ways. First, related to cognitive skills, the inter-rater reliability among judges without training is notoriously low in evaluation of such works as student essays. This is also true for one judge who evaluates the same work after the passage of time.[47] Unless librarians are carefully sensitized to the issues involved, I suspect we would vary widely as to judgments from our reading of students' journals and paper regarding the degree of successful application of a library search process.

RELIABILITY

IS IT STABLE AND CONSISTENT

Second, regarding attitudes, measures of library-related attitudes often may not appear reliable simply because these attitudes are not stable. Various studies have shown that students' attitudes toward their college experience tend not to stabilize until well into their sophomore year. If we evaluate the attitudes of freshman students towards the library, we may be simply evaluating their attitude towards their most recent experience with the library. Did they find the last book for which they looked? A reliable measure for one purpose with one group may not be reliable for another group for even a closely related purpose.

There are different kinds of reliability, which can be determine in different ways. The test-retest method involves administration of the same test after a period of time and correlation of the results. As simple as this may seem, it has several practical and theoretical limitations. For example, memory becomes a systematic source of variance and a substantial but misleading correlation can result. On the other hand, over a period of time the characteristics being measured might themselves change, and the test-retest method in this case could result in an underestimation of the reliability of the measure.

The difficulties encountered in the test-retest method only partially can be avoided through the use of the equivalent-forms method. With this method the same individuals are tested with one form of a test and then tested with an equivalent form. In addition to sharing some of the problems of the test-retest method, this method adds the problem of determining if a truly equivalent form is being used.[48]

Estimates of internal consistency reliabilty can be determined by the split-half method or related methods. The split-half method involves simply dividing the test in half or on an odd-even or similar basis and computation of the reliability coefficient using the scores from the halves. A problem with this method is that the correlations between the halves will vary depending on how the items are divided. There are statistical packages, such as RELIABILITY in SPSS, that overcome this limitation, and it is this reliability coefficient you see most often reported in the literature. (overhead transparency off)

Conclusion and Recommendations

All these limitations and considerations regarding evaluation can be discouraging, but they are simply a reflection that we live in a complex and less than perfect world. We are being evaluated in our efforts by those from we seek acceptance and support whether we want this evaluation or not. It is to our benefit to make this evaluation as sound and as rigorous as possible. Certainly the history of bibliographic instruction has shown that the acceptance of it as an integral part of higher education is not inevitable. Despite the popular belief that we live in an age of unprecedented change, securing the acceptance of a new idea or the adoption of an innovation often is difficult. In the commercial marketplace, it has been estimated that only one idea out of more than every five hundred results in a successful new product.[49] I suspect that given the conservative nature of higher education, the acceptance rate in this arena is even much less.

Education contains many examples of the struggles involved in gaining the acceptance of an innovation. Kindergarten is just one example. After its introduction into the United States before the turn of the century, more than fifty years lapsed before it was widely adopted during the 1930s and 1940s. In fact, it has been estimated that public education has a fifty-year time lag. A popular prediction based on past history is that it will be approximately the year 2000 before the current movement of educational reform started after World War II will be successful.[50] Even with the educational crisis created by Sputnik many school systems remain virtually untouched by the huge federal and state expenditures that promoted educational reform during the past thirty years.

I think we have recognized the work needed to promote bibliographic instruction, and we are exploring several avenues. In addition to formative and summative evaluation, we are trying to gain inroads with other disciplines through their professional societies. Obviously, we have a long way to go. I will end this paper as I started it with reference to a quote attributed to Mark Twin. Recently I have been involved in efforts to weed materials from the library I direct. Inevitably someone will select a book and say, "We can't weed this one, it is a classic." It is times like this when Mark Twain's definition of a "classic" comes to mind. A classic, according to Twain, is, "A book which people praise and don't read." I suspect as with classic books, we librarians do more praising of the evaluation of bibliographic instruction than doing it. I hope that I have given you some encouragement today to evaluate your efforts.

REFERENCES

1. Charles Dudley Warner first wrote in an editorial in the Hartford Courant in 1890 that "everybody talks about the weather, but nobody does anything about it." Even then it was so often attributed to Mark Twain that Charles Hopkins Clark, editor of the Courant wrote: I guess it's no use; they still believe Mark Twain said it, despite all my assurances that it was Warner.

2. Richard Hume Werking, "Evaluating Bibliographic Education: A Review and Critique," Library Trends 29 (Summer 1980): 153-171.

3. Association of College and Research Libraries, Bibliographic Instruction Section Research Committee Subcommittee on Evaluation, Evaluating Bibliographic Instruction: A Handbook (Chicago: American Library Association, 1983).

4. Kenneth Brough, Scholar's Workshop: Evolving Conceptions of Library Service (Urbana, Illinois: University of Illinois Press, 1953), p. 152.

5. John Mark Tucker, "Emerson's Library Legacy: Concepts of Bibliographic Instruction," in Increasing the Teaching Role of Academic Libraries, ed. Thomas G. Kirk (San Francisco: Jossey-Bass, 1984), pp. 15-24.

6. Peter Hernon, "Instruction in the Use of Academic Libraries: A Preliminary Study of the Early Years as Based on Selective Extant Materials," Journal of Library History 17 (Winter 1982): 16-38.

7. Harvie Branscomb, Teaching With Books (Chicago: Association of American Colleges, American Library Association, 1940).

8. Ibid., p. 37.

9. Louis Round Wilson, Mildred Hawksworth Lowell, and Sarah Rebecca Reed, The Library in College Instruction (New York: H. W. Wilson, 1951), p. 13.

10. Patricia B. Knapp, "A Suggested Program of College Instruction in the Use of the Library," Library Quarterly XXVI (July 1956): 224-231.

11. Patricia B. Knapp, The Monteith College Library Experiment (New York: The Scarecrow Press, 1966).

12. James R. Kennedy, Jr., "Integrated Library Instruction," Library Journal 95 (April 15, 1970): 1450-1453.

13. Larry L. Hardesty, John P. Schmitt, and John Mark Tucker, eds., User Instruction in Academic Libraries: A Century of Selected Readings (Metuchen, N. J.: The Scarecrow Press, 1986), p. 228.

14. Gresham Riley, "Myths and Realities: The Academic Viewpoint," in Academic Libraries: Myths and Realities; Proceedings of the Third National Conference of the College and Research Libraries, edited by Suzanne C. Dodson and Gary L. Menges (Chicago: Association of College and Research Libraries, 1984), p. 12.

15. Robert N. Broadus, "Use Studies of Library Collections," Library Resources & Technical Services 24 (Fall 1980): 317.

16. Edward A. Suchman, Evaluative Research (New York: Russell Sage Foundation, 1967), p. 2.

17. Ibid.

18. Larry Hardesty and Frances Gatz, "Application of Instructional Development to Mediated Library Instruction," Drexel Library Quarterly 16 (January 1980): 3-26.

19. Robert M. Diamond, "Piecing Together the Media Selection Jigsaw," Audiovisual Instruction 22 (January 1977): 50-52.

20. Larry Hardesty, Nicholas P. Lovrich, Jr., and James Mannon, "Evaluating Library-Use Instruction" College and Research Libraries 40 (July 1979): 309-317; Larry Hardesty, Nicholas P. Lovrich, Jr., and James Mannon, "Library-Use Instruction: Assessment of the Long-Term Effects," College and Research Libraries 43 (January 1982): 38-46.

21. Constance Mellon, "Library Anxiety: A Grounded Theory and Its Development," College and Research Libraries 47 (March 1986): 160-165.

22. Benjamin S. Bloom, ed., Taxonomy of Educational Objectives: Handbook I: Cognitive Domain (New York: David McKay Co., Inc., 1956), p. 7.

23. Masse Bloomfield, "Testing for Library-Use Competence," in Educating the Library User, ed. John Lubans, Jr. (New York: R. R. Bowker, 1974), p. 230.

24. Thomas Kirk, "Evaluation of Library Orientation and Instruction Programs: A Taxonomy," in Planning and Developing a Library Orientation Program (Library Orientation Series No. 3), ed. Mary Bolner (Ann Arbor, Pierian Press, 1975), p. 47.

25. Abraham Kaplan, The Conduct of Inquiry: Methodology for Behavioral Science (San Francisco: Chandler, 1964), pp. 3-11.

26. Bloom, Taxonomy.

27. Ibid., p. 63.

28. Ibid., p. 89.

29. Ibid., p. 120.

30. Ibid., p. 144.

31. Ibid., p. 162.

32. William G. Perry, Jr., Forms of Intellectual and Ethical Development in the College Years (New York: Holt, Rinehart and Winston, 1968).

33. Ibid., p. 9.

34. Bloom, Taxonomy, p. 185.

35. Scarvia B. Anderson, Samuel Ball, Richard T. Murphy and Associates, Encyclopedia of Educational Evaluation (San Francisco: Jossey-Bass, Inc., 1973), p. 191.

36. Ibid.

37. Hardesty, Lovrich, and Mannon.

38. Werking, "Evaluating Bibliographic Education."

39. Anderson, Encyclopedia of Educational Evaluation, p. 191.

40. Knapp, Monteith College.

41. Anderson, Encyclopedia of Educational Evaluation, p. 192.

42. Werking, "Evaluating Bibliographic Education," p. 166.

43. Anderson, Encyclopedia of Educational Evaluation, p. 191.

44. Committee to Develop Standards for Educational and Psychological Testing of The American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education, Standards for Educational and Psychological Testing (Washington, D.C.: American Psychological Association, Inc. 1985).

45. Bonnie G. Gratch, "Research Designs Appropriate for Evaluating Bibliograhpic Instruction," in Evaluating Bibliographic Instruction: A Handbook, pp. 37-60.

46. Ibid., p. 47.

47. Charles R. Cooper and Lee Odell, eds., Evaluating Writing: Describing, Measuring, Judging (Urbana, Illinois: National Council of Teachers of English, 1977), p. xi.

48. Anne Anastasi, Psychological Testing, 5th ed. (New York: Macmillan Publishing Co., 1982), pp. 112-113

49. Elizabeth Martin, New Products, New Profits (New York: American Management Association, 1964), p. 9, as cited by Everett M. Rogers, Communication of Innovation (New York: The Free Press, 1971), p. 68.

50. Don Glines, "Why Innovative Schools Don't Remain Innovative," The Education Digest XXXVIII (May 1973): 5.